Extract Agent-based Model from Social Communication Network

Hung-Ching (Justin) Chen
chen3@cs.rpi.edu

Matthew Francisco
francm@rpi.edu

Malik Magdon-Ismail
magdon@cs.rpi.edu

Mark Goldberg
goldberg@cs.rpi.edu

William A. Wallace
wallaw@rpi.edu

June 1, 2007

Abstract

We present a machine learning methodology (models, algorithms, and experimental data) to discovering the agent dynamics that drives the evolution of the social groups in a community. There are a number of challenges, the first and foremost being the complex nature of the *micro-laws*, determining the behaviors of an agent, needed to represent even a very simple society – agents may have discrete attributes together with continuous parameters, which inevitably leads to a mixed optimization problem, and each agent has its own attributes, which also interact with others attributes, suffering from combinatorial and dimensionality curses. Another challenge is that the data upon which to answer the question is not available – typically social groups (especially online groups) do not announce their membership and one has to infer groups from observable macro-quantities such as communication statistics. We set up the problem by introducing a parameterized probabilistic model of social group evolution, called ViSAGE (Virtual Simulation and Analysis of Group Evolution), for the agent dynamics: the actions of an agent are determined by *micro-laws* with unknown parameters. Our approach is to identify the appropriate *micro-laws* corresponds to identifying the appropriate parameters in the model. To solve the problem we develop heuristic expectation-maximization style algorithms for determining the *micro-laws* of a community based on either the observed social group evolution, or even observed set of communications between actors, the most of cases in a real community. In order to avoid the resulting combinatorial explosion, we appropriately approximate and optimize the objective within a coordinate-wise gradient ascent (search) setting for continuous (discrete) variables. Finally, to test the quality of our approximations and the feasibility of the approach, we present the results of extensive experiments on simulated data as well as some results on real data, *e.g.*, newsgroups, emails, and blogs.